

## Weekly Report

### Done

1. 思路有点乱,可能有点过于囿于彭老师的文章了.周末绞尽脑汁没有想的太清楚.现在我从技术角度\可视化任务角度\行为分析对象角度进行思考,他们构成一个设计空间.我明晚或者后晚先再和彭老师聊聊.
2. 讨论:
  - i 咨询了谢潇一些 embedding 技术,不过他做的大多数仍然在动态图 embedding
  - ii 和梅博交流一些他的想法.时间上可以聚合(切分一周的工作日和周末,一天不同时段);找到代表性的用户(比如我们把 event 变成点.假设有一类用户,我们把他们所经历的 event 连成线,这样就会有很多线,然后用某种技术找到最能代表整体趋势的线作为一个代表);用户分类:游戏开新服时,有些用户是老玩家过来,有些是新手;尽管做着类似的任务,但可能有不一样模式,这对 embedding 可能也有影响.不过这一点正向看很容易想通,比如我已知两类人,他们分别各自做 embedding,各自结果中一个行为周围其他行为分布可能有很大不同;但是如果我把两类人都一起 embedding 了,我怎么发现这种情况?有待思考.
  - iii 咨询了一下李权,他已在美国交流时做过一些关于游戏中的动作(以等级作为时间)的 event2vec.在这方面,光从投影角度,他没有发现很有用的 insight;他指出在游戏设计时候,动作的出现不会因为等级变化有太大变化.这是一个帮我们避开的点;当然我们拥有交流信息/消费信息,这些信息相当于动作而言应该会有一些更丰富的变化在里面.他建议,还是需要找一些在不同的阶段语义信息变化较大的数据,也方便做展示;此外,他今年的 embeddingvis 论文,其实是强行的打破 embedding 与可解释性的鸿沟;其实是很有意思的.假设我们可以通过不同的时间/session 划分方式来做 embedding,对于相同的行为/时间是否也可以出现不同的 pattern?然后我们也可以结合具体信息来分析原因.
3. 修改 PVis 论文细节以及视频内容.
4. 对相关 embedding 技术进一步了解,对 kdd18 文章的联合嵌入部分进一步咨询了知友,可以在下周尝试.
5. 一些数据测试:时间序列变网络,在网络上找 pattern,特征化,用于分类(未完成).

### To Do

1. 整理思路,与各位老师同学讨论,敲定论文方向.
2. 测试上述未完成的数据模型;整理游戏中的行为数据.

### 论文阅读

1. **KDD18** Sequence of sets 本文不在对时间序列按一个个 item 进行预测,而是变成一个个 set 预测;举个例子就是你每次发邮件给一批人,这些人构成一个 set,你会有一系列 sets.根据一

些非常简单的参数来预测, 参数估计主要用到极大似然估计. 这也是一种类似将 individual 的 item 变成 session 来做分析预测的方法.

2. Mobile phone use as sequential processes 彭老师那边的文章研究的两大 sequential processes 是 mobile session 和 mobile trajectory. 这里提供了研究的不同维度, 尤其是 inter-individual 和 intra-individual 等, re-engagement 和 behavioral complexity 等. 其实如果光展现这些信息也有很大可视化工作量, 但是 novelty 方面就有待考量; 我觉得需要引入一个可交互的元素, 来辅助这样分析与展示.
3. EventThread1 和 2, 分别是去年今年 VIS. 文章很好的跨越了对于事件序列, 数据分析的粗粒度和可视化细粒度的鸿沟; 并在 2 中应用 embedding 提高了时间切分的自由度和整个 stage 分割的自动化程度. 这里面用到的 embedding 技巧, 容入了事件间隔的时间长度, 可以借鉴学习.
4. EmbeddingVis 在上面提到了其优点. 此外其中提到的 NDCG(Discounted cumulative gain)用于评估搜索结果的指标(相似性与搜索结果靠前)也是一个很好用的度量方法.

## 工作时间

平时 50, 周末一共 16 小时. 总共 66 小时.